Missouri Department of Health Cancer Inquiry Protocol

A manual for reporting cancer clusters in Missouri communities

Missouri Department of Health
Division of Chronic Disease Prevention
and Health Promotion

March 1999

Cancer Inquiry Protocol

Prepared by:

Eduardo J. Simoes, M.D., M.Sc., D.H.L.S.T.M., M.P.H.
Chronic Disease Medical Epidemiologist

First Edition

Suggested Citation

Missouri Department of Health
Cancer Inquiry Protocol:
A manual for reporting cancer clusters
in Missouri communities

Ordering Information

To order a copy of this manual,
Call the Missouri Department of Health,
Bureau of Cancer Control,
At 573/522-2845 or write to the following address:

Missouri Department of Health
Bureau of Cancer Control
P.O. Box 570
Jefferson City, MO 65102

Acknowledgements

This manual was prepared by the Missouri Department of Health, Division of Chronic Disease Prevention and Health Promotion.

We thank the Cancer Inquiry Program and the Cancer Inquiry Committee who discussed early ideas for this manual with us, reviewed drafts and made recommendations on format and content.

We particularly would like to acknowledge the assistance of the following persons.

For providing information and comments:

Ross Brownson, Ph.D.
St. Louis University
Marianne Ronan, M.P.A.
Bureau of Cancer Control
Sarah Arbuckle
Cancer Inquiry Program
Philip N. Brunner, M.S.
Center for Local Public Health Services
Scott Clardy,
Section for Environmental Public Health
Jian Chang, Ph.D.
Missouri Cancer Registry
Robert Hamm, M.D., M.P.H.
Office of Epidemiology
Jeannette Jackson-Thompson, M.S.P.H., Ph.D.
Office of Surveillance, Research and Evaluation
Scott Clardy
Section for Environmental Public Health
Wayne Schramm, M.A.
Bureau of Health Data Analysis
Margaret Tyler
Bureau of Health Data Analysis
Michael Carter, M.H.A.
Bureau of Health Promotion
Bernard Malone, M.P.A.
Missouri Department of Health
Sherri Homan, R.N., M.S.N., C.F.N.P.
Division of Chronic Disease Prevention and Health Promotion
Larry Erickson
Missouri Department of Natural Resources
Joe Leslie
Missouri Department of Agriculture

Kathleen Anger, Ph.D.
Horizon Research Services
John Gulick
Community Health Assesment Resource Team
Jeanie Shaneberger
Breast and Cervical Cancer Control Project
Gary Wilson,
Bureau of Health Promotion
Denise Jordon-Izaguirre
Agency for Toxic Substances and Disease Registry
Sue Vest, CTR
Missouri Cancer Registry
Alyce Turner
Citizen Representative

For text design and prepartion:
James McDonald, M.A.
Mary Kay Hager

For clerical support:
Jamie Strawn

Table of Contents

TABLE OF CONTENTS (continued)

Introduction

The systematic approach in decision-making outlined in this protocol is based on the basic principles of epidemiological reasoning and causation. This protocol aims to assure a greater likelihood of usefulness and cost-effectiveness for the cancer inquiry process. Cancer inquiries go through at least the first level of a four-level process, as described in the following pages. Each step uses epidemiological reasoning and/or analysis to determine if the inquiry should proceed to the next level.

At all levels, a report with recommendations for action for each cancer inquiry is reviewed by the Cancer Inquiry Committee, an interdisciplinary committee of epidemiologists, environmental specialists, health educators, statisticians, other specialists and citizen volunteers (see *Missouri Cancer Control Program* for a more complete description). These recommendations are prepared by the Chronic Disease Medical Epidemiologist and the Cancer Inquiry staff.

The Cancer Inquiry Process

Initiation of the Cancer Inquiry Process

The Cancer Inquiry (CI) process begins with contact from a citizen, a health sector representative (local public health administrator, physician, nurse, certified tumor registrar, etc.), the Environmental Protection Agency, or a CI Committee member regarding perceived excess cancer cases or cancer-related deaths in a community, neighborhood, or other defined area or sub-population.

If a citizen contacts the CI Program for a purpose other than a concern about excess cancer occurrences, then the CI staff will send the citizen an answer to the question(s) asked. This is classified as a cancer *concern*. An example would be a mother calling for information on Burkitt's lymphoma because her son was recently diagnosed. She is not concerned about an excess of cancer occurrence. (Appendix A)

If the citizen who contacts the CI Program is concerned about a geographical area studied in another inquiry, the final report related to the area of interest is sent to the citizen and the inquiry is classified as a cancer *concern*.

Perceived cancer incidence and/or mortality excesses in a defined area or sub-population, over time, or compared to that from other areas are classified as cancer *inquiries*.

LEVEL 1 Inquiry

The Level 1 Inquiry begins when a citizen contacts the CI Program about an excess of cancer. A member of the CI staff conducts a lengthy telephone interview with the inquirer, usually taking about an hour. During this interview, the staff member collects information necessary to complete the CI Initial Report Form (Appendix B), including identifying information regarding the inquirer (e.g., the inquirer's name, address and phone number) and basic information about the citizen's concern. The inquirer is asked to provide general information about the number of excess cases, the types/sites involved and suspected exposures or hazards that may be contributing to the perceived excess.

The remainder of the interview focuses on education about cancer and the CI Program and on giving the inquirer information on the prevalence, risk factors, and screening opportunities for cancer. The staff member will also briefly explain the next steps of the inquiry process.

All correspondence with inquirers is copied to the Section for Environmental Public Health (SEPH), the appropriate Missouri Department of Health (DOH) District Health Director, the DOH Center for Local Public Health Services, the appropriate local

public health agency, and any other involved parties. However, the name of the inquirer is kept confidential.

After the interview, the staff member assigns the inquiry a case number. This case number consists of the year in which the first contact occurs and a number assigned in order (e.g., the fifth call received in 1998 would be 98-005.) The First Response Letter (Appendix C) is prepared and sent to the inquirer along with other relevant information:

- *Cancer Facts and Figures* from the American Cancer Society
- Cancer Excesses Fact Sheet (Appendix D)
- Risk factor information (Appendix E)
- Brochures on the specific type(s) of cancer of concern (e.g.,lung cancer, breast cancer, etc.)
- Mortality and/or incidence rates (if applicable)
- Any other information specific to the concern

This packet also includes the Response Form and Patient Listing Forms (Appendix F), which were explained to the inquirer during the phone interview. The inquirer has six weeks to return the forms or contact the CI staff to notify them that additional time is needed to obtain the information requested.

The Response Form consists of a series of questions regarding the case identification and definition as well as identification of the potential exposure. The Patient Listing Form asks for detailed information about each case of cancer suspected to be related to the exposure. The inquirer is requested to provide a brief description of each case, including date of diagnosis, smoking history, occupation, and the number of years the individual resided in the area before the diagnosis was made. If the inquirer does not contact the CI staff or return the Response Form and Patient Listing Forms within the six weeks, the inquiry is considered closed and a certified, return-receipt letter is sent (Appendix G) explaining the outcome.

If any inquiry involves issues that are of concern to other local, state, or federal agencies, such as the Department of Natural Resources (DNR), the Department of Agriculture (DOA), or SEPH's Health Consultation Team, the issue is referred to the appropriate agency.

Because cancer incidence data submitted to and maintained by the Missouri Cancer Registry (MCR) is not coded in a way that easily allows researchers to track occupational clusters, cancer concerns that center on a worksite may be referred to either the National Institute for Occupational Safety and Health (NIOSH) or to the Occupational Safety and Health Agency (OSHA). Because NIOSH's procedures require that an employee or manager submit the complaint, the application for the process will be included in a

letter to the inquirer. However, the CI Program can refer concerns directly to OSHA if there is evidence that a public health threat may be present. In either instance, DOH will assist in any research conducted as a result of a referral.

If the inquirer returns the Patient Listing and Response Forms, the CI staff sends a Confirmation Letter (Appendix H) which explains more about the CI process and informs the inquirer when the next meeting of the CI Committee will be held. Then CI staff, in consultation with the Chronic Disease Medical Epidemiologist and the Chief of the Bureau of Cancer Control (BCC), reviews the forms and determines if the inquiry meets the criteria to conduct further research.

Case verification and ascertainment are initiated in all cases where patient listings are received. Case verification is the process of verifying the accuracy of information on cases listed by the inquirer. Case ascertainment involves identification of new cases through an active search of existing databases. Information obtained through case verification should include the site, histology and stage of the cancer at diagnosis, date of diagnosis, gender, race, smoking status, occupation, date of birth, and address at diagnosis of each cancer case. The MCR database is used for most case verification. Doctors' records, hospital records, and/or death data are alternative sources for case verification. Cases not confirmed by a medically reliable source are excluded from data analysis.

The following conditions must be met in order to consider pursuing a Level 2 Inquiry:

1. Case verification is completed.
2. A definable type of cancer is reported.
3. Only one type of cancer is reported or the types of cancer reported have common suspected risk factors (based upon the existing database of risk factor literature searches) including exposures that have been reported to be a possible cause of multiple tumors (e.g., dioxin and radiation).
4. A specific environmental or occupational cause (exposure) is suggested for this excess.
5. There is a plausible scenario for the patients to have come into contact with the suspected cause of their cancer.
6. There is no evidence of common behavioral-type or other risk factors with strong, well-proven relationships to the identified cancer, such as between smoking and lung cancer; between polyvinyl exposure in occupational settings and angiosarcoma; and between asbestos and mesothelioma.

In addition, at least two of the following characteristics must also be present:

1. A common type of cancer is occurring in an unexpected age group.
2. The cancer of interest is rare.
3. The suspected exposure is plausibly linked to the cancer(s) of concern, based on the available knowledge base and research.
4. If the cancer latency in relation to a particular exposure is known, there is a reasonable match between the estimated time cases have been exposed and this latency. Unless the cluster is in an occupational setting, the time a case has been exposed is estimated by the time a case lived in the area of concern and the time exposure was first established. If the cluster is in an occupational setting, the time of exposure is estimated by the time the case was employed and when exposure was first established. For most cancer sites, if cancer latency is not known, a minimum of ten years' residency in the study area or exposure in the occupational setting is necessary for the suspected association to be considered plausible. Childhood cancers may have a shorter latency period.

In many situations, a literature search will not support the plausibility of the association between the exposure and the cancer of concern. However, it is always possible that a new and not yet studied association, including a different pathway for exposure, may exist.

If the Response and Patient Listing Forms are not returned and no further contact is made by the inquirer within the allotted time, there are two instances in which the CI Program can still proceed to Level 2:

- The Chronic Disease Medical Epidemiologist, in collaboration with CI staff, concludes that there is enough information to indicate a potential public health risk.
- A situation involves heightened community concern and DOH administration or another DOH division determines such an action is warranted.

LEVEL 2 Inquiry

A Level 2 Inquiry consists of an epidemiological evaluation by the Chronic Disease Medical Epidemiologist and the MCR senior researcher to determine if there is a cluster.

During a Level 2 Inquiry, an epidemiological evaluation is conducted. The analytical research conducted should include:

1. Literature searches for additional information about risk factors.
2. Calculations and statistical testing by geographical area, time periods, and major demographic groups.

If clustering of cancer occurrence is supported by the Level 2 research, the following conditions must be met before moving to a Level 3 Inquiry:

1. Case ascertainment, generally initiated in Level 1, is completed.
2. All conditions and characteristics set forth as criteria for moving to a Level 2 Inquiry still hold.
3. There must be at least five cases of one type or related types of adult cancer or at least three cases of one type or related types of childhood cancer per sub-population of interest (age group within gender and race) in the study area.
4. Statistical calculations indicate the finding of a cancer cluster. Supportive findings include: statistical results are consistent across tests (congruency among spatial clustering testing and congruency among temporal clustering testing) and robust to statistical flaws intrinsic in the tests (assumptions are met) and p-values are highly significant (≤ 0.001).[1]
5. There must be a population large enough and well-defined enough to conduct a more detailed analytical study with further classification of sub-populations.

After reviewing the Chronic Disease Medical Epidemiologist's recommendations and other relevant information, the CI Committee may determine that there is no cancer cluster in the community, in which case the inquiry is closed. However, if there is any epidemiological evidence of a cancer cluster in the community, and the conditions outlined above are met, the inquiry moves to Level 3. In either instance, a letter is sent to the inquirer explaining the reason for closing the inquiry or proceeding with additional analyses. while providing any other relevant findings.

A Level 3 Inquiry may also be pursued without the above conditions being met if, in the opinion of the CI Committee, other aspects of the inquiry justify further analytical epidemiological study of the association between the suspected exposure and the cancer of interest. The CI Program may also pursue a Level 3 Inquiry if the CI Committee and DOH administration find that a heightened community concern is still present for inquiries initiated out of this concern.

LEVEL 3 Inquiry

During a Level 3 Inquiry, a more detailed and in-depth epidemiological evaluation is conducted. Exactly what additional research is conducted may vary, but can include any or all of the following:

- Literature searches for additional information about risk factors;
- Additional case ascertainment that encompasses a wider geographical area, additional time periods, and/or searches of other databases;
- Data analysis incorporating either further defined demographic sub-groups, a wider geographical area, or additional time periods;
- Additional statistical testing for assessing consistency in defining the cluster;
- Collection of additional information on the environmental concern, which may include cooperation with other agencies or departments;
- Other information relevant to the specific inquiry.

Once a Level 3 Inquiry is completed, it will be presented to the CI Committee.

It is recommended that a Level 4 Inquiry should be pursued only if the following conditions are identified at the end of a Level 3 Inquiry:

1. All conditions and characteristics set forth as criteria for moving to Levels 2 and 3 still hold.
2. There is a population large enough and well-defined enough to conduct a more detailed analytical study.
3. The epidemiological investigation and statistical testing unequivocally indicate clustering.

The following conditions are necessary for statistical testing to unequivocally indicate clustering:

1. Additional calculations using related alternative time and space clustering statistics are consistent among themselves and with previous statistics.
2. Additional calculations for time and space clustering find a highly statistically significant increase of cancer incidence over one time period or a significant differential between areas (P-value ≤ 0.001; one-sided).[1]
3. Additional calculations for time clustering find a statistically significant increase (P-value ≤ 0.025; one-sided) over two consecutive time periods.

A Level 4 Inquiry may also be pursued if, in the opinion of the CI Committee, other identified aspects of the inquiry warrant the pursuit of an analytical epidemiological study of the association between the suspected exposure and the cancer of interest. Also, the CI Program may pursue a Level 4 Inquiry if the CI Committee and the administration of DOH conclude that a heightened community concern is still present for inquiries initiated out of this concern.

If clustering is determined at Level 3, an environmental assessment may be initiated or completed before proceeding to Level 4. This assessment could be conducted initially by the Missouri Department of Natural Resources with the help of federal environmental agencies. If a current exposure or an environmental or occupational health hazard is identified, efforts should be initiated to control or prevent further exposure.

LEVEL 4 Inquiry

A Level 4 Inquiry consists of an initial feasibility study and design of an analytical study. It may also consist of an environmental assessment followed by a feasibility study, if an assessment was not previously completed. Also, at this level if an analytical study is feasible, additional case ascertainment should be pursued and completed.

The feasibility study is designed to determine if an epidemiological investigation should be conducted. Complete case ascertainment, identification of the requisite minimum number of cases in an identifiable study population of sufficient size, and the ability to measure environmental exposure both at the community and at the individual level are the critical elements in determining the feasibility of an analytical study. The researcher should also evaluate the feasibility of an analytical study when either the suspected exposure or cancer occurrence is rare.

DOH may ask outside agencies, such as the Centers For Disease Control and Prevention (CDC) and the Agency for Toxic Substances and Disease Registry (ATSDR), to participate in the feasibility study, the additional environmental exposure assessment, or the design and implementation of the analytical study. This generally occurs when a cancer cluster has been found and the resources needed to conduct further research are beyond the capacity of DOH.

Outside researchers must submit a research plan and receive Institutional Review Board (IRB) approval before confidential data is released.

Analytical Methods for Cancer Inquiries

Epidemiological Evaluation of Clustering

The evaluation of cancer clusters should not be different than other epidemiological evaluations with respect to the use of epidemiological reasoning and principles. In the same manner of analytical studies, the consistency, biological plausibility, and dose-response of the suspected disease and exposure association should be evaluated. These principles should be used to guide the Chronic Disease Medical Epidemiologist and the CI staff in every step of the process to determine if there is a cancer cluster and if there is a possible cause for that cluster that should be examined in an epidemiological study.

The criteria for deciding whether to initiate analytical work at each level of the cancer inquiry should assess a number of issues: the case and exposure definition (ill-defined case or exposure); latency between initial exposure and diagnosis of disease; plausibility of the exposure pathway; and, confounding and other possible explanations for the observed increased incidence and mortality. After conclusion of a Level 3 Inquiry, the feasibility of an epidemiological investigation (Level 4 Inquiry) should be questioned if the number of cases is small, cancer occurrence is rare, the exposure is rare, both cancer occurrence and exposure are rare, or the ability to measure exposure at both community and individual levels is weak.

This process should also be conducted with the realization of the limitations for establishing causal links using either surveillance or individual level data that is incomplete. The evaluation of cancer clusters involves the use of exploratory methods capable of generating, not testing, hypotheses.

In addition, the limitations of statistical testing in epidemiological settings should be considered. Statistical testing at its best can only assist causal reasoning. The "significance" of the epidemiological causality principles and the "significance" of patterns and trends of health events and their implications for public health far outweigh the "significance" of a P-value (probability of a type-I error).[5]

Finally, whenever possible, there should be no place for ambiguity in epidemiological evaluation. The p-value is an ambiguous measure that is both a function of the uncertainty (precision of the measure) of observing the effect being evaluated and the magnitude of the effect. Given a large enough sample, a "significant" effect or an association can be demonstrated even when it is small and with little epidemiological or public health significance. For this reason, it is good epidemiology to separate out the two components of an association in a meaningful way

when presenting epidemiological results. For example, rate ratios and standard error of the ratio with subsequent calculation of a 95% confidence interval on the ratio will do exactly that. The rates and rate ratios provide information on the effect and its magnitude; the standard error provides information on the precision of the estimate.

As an illustration for this discussion, consider the following scenario: a call is received from a concerned citizen about a cluster of six cancer cases in one year in a single neighborhood (a city block). The cancer cases reported include breast, colon, lung, uterine, skin, and prostate cancer among individuals aged 65 and older. Using epidemiological reasoning and criteria for causation, these cases could not be environmentally related, despite the demonstration of a statistically significant cluster. On the other hand, consider also another scenario: a call is received from a concerned citizen about a cluster of 16 cancer cases in one year in a single neighborhood (a zip code) and possible dumping of hazardous material by an industrial plant. The cases reported are all lymphatic cancer among individuals aged 65 and older. The difference between observed cases and expected cases is not significant, and could have occurred by chance. Using epidemiological reasoning and criteria for causation, these cases could well be environmentally related, despite the demonstration of a statistically non-significant cluster.

Study Area

For all levels of a cancer inquiry, the study area is the smallest detectable geographical area, using the most recent population database (US Census and the Center for Health Information Management and Epidemiology's [CHIME] annual updated estimates) from which cases arise that allows for reliable calculations of rates and statistical testing. These areas may include one or more of the following: county, zip code (ZIP), census tract (CT), Standardized Metropolitan Area (SMA), or city. A SMA is an urban area with a minimum of 50,000 inhabitants living within defined geopolitical boundaries, as defined in the 1990 US Census.

For the purpose of rate calculation, a city and a SMA may or may not overlap the same geographical area. For directly adjusted rates, if a SMA contains a city with at least 50,000 inhabitants and a sufficient number of cases arise from the city, calculations will be implemented using the city as the study area.[2] If a sufficient number of cases arise from a city with a population smaller than 50,000 but within a defined SMA that meets the above conditions, the SMA will be defined as the study area. For indirectly adjusted rates, the same rationale is used with an area population of at least 10,000.[1] If cancer cases arise from other geopolitical areas (ZIP

code or county) for which both observed cancer cases and population size are appropriate, calculation of rates and statistical testing will use these as the study area.

Other possible study areas may include schools and occupational settings for which both the total number of cases and the total population size can be reported and later verified and ascertained.

Typically, areas larger than and contiguous to the study area should be used as comparison during a cancer cluster investigation.

It is also possible that after case verification (Level 1 Inquiry) and ascertainment (Level 2 Inquiry), the boundaries of the study area may be modified.

Study Period

Usually, the study period corresponds to that of identified incident or deceased cases reported on the Response and Patient Listing Forms for all inquiries at Level 2 and above. The study period may correspond to that of verified and ascertained incident or deceased cases for a Level 3 Inquiry. In a situation where an insufficient number of cases is identified but an inquiry is perceived as necessary, the CI Committee, in consultation with the Chronic Disease Medical Epidemiologist, may decide to pursue further research. If available, more years of incidence and mortality data are used than those indicated by the Response and Patient Listing Forms. Also, when the evaluation of evidence for clustering at Level 2, Level 3, or Level 4 is not statistically significant, but there is plausibility for establishing a cluster, more years may be used to increase the statistical power of the analysis.

In either situation, the ideal denominator for rate calculations is that of population estimates of corresponding years from which verified and ascertained cases arise.[3] In practice, depending on the definition of the study area, if population-based data by demographic groups is not available, the population estimates of the last US Census are used for all yearly and aggregate estimates of rates. A historical series of either incident or deceased cases in a study area which are available in the DOH database may also be used to evaluate randomness of occurrence over time and space.

Population

A choice of study (population at risk or index population) and comparison populations for statistical calculations, including area and period as well as sub-populations (gender, race, and age groups), will be dictated by the availability of population-based data.

Population-based data used for calculation of rates may include the US 1990 Census, the Missouri 1990 Census, or updated Missouri inter-census population estimates by gender, race, and age group provided by the National Center for Health Statistics (NCHS) and CHIME, respectively. For standard US and Missouri comparison rates, population-based rates are either the Surveillance, Epidemiology, and End Results (SEER) Program or the Missouri Cancer Registry (MCR), respectively, depending on the time period involved.

The comparison population during the period of concern for calculations of directly standardized incidence and mortality ratios will be the US and Missouri for incidence (SEER and MCR incidence estimates) and Missouri for mortality (CHIME mortality estimates), respectively. The US 1970 Census Standard Population (Million) is the standard used for direct age-adjustment of rates. The comparison populations from which standard rates are used for indirect age-adjustment during the same period of concern are either the population of the US (SEER incidence) or Missouri (MCR incidence and CHIME mortality). The study population is the index population as defined using geographical and time period criteria described above.

Ideally, at Level 2, sub-populations should include classification of cancer occurrence and death by gender and race regardless of the choice for study area and period. At Level 3 and Level 4, further classification of rates into age-specific groups and possibly by occupation and smoking status is warranted, if possible. An exception to this classification of sub-populations is childhood cancers, for which no further separation by gender is attempted. Also, because index (study) population observed rates and counts are compared to rates expected from a comparison population, the choice of the study population is only made after identification of a comparable population.

Cancer Sites

Only primary sites are studied using International Classification of Diseases-9 (ICD-9) codes.[4] For initiation of Level 2 Inquiry calculations, it may be assumed that the case notification by the citizen on the Response and Patient Listing Forms is adequate. Whenever possible, calculations should be made using this information and additional cases (sites) verified by MCR for the study area. For Level 3 and Level 4 Inquiries, thorough case ascertainment of diagnosis will determine accuracy of reported sites.

If the citizen's report involves more than one site, attempts will be made to calculate incidence and mortality statistics for all sites involved during a Level 2 or Level 3 Inquiry. If only one cancer site is involved, then incidence and mortality statistics are calculated for this site and "all cancer" by the same sub-populations as above. If the inquiry concerns childhood cancer, the SEER pediatric major cancer group that includes the cancer of interest will be used instead.[3]

Case Verification and Ascertainment

Case Verification

For a Level 2 Inquiry, information from the Response Form and Patient Listing Forms is verified for accuracy of diagnosis (site, histology, and stage), date of diagnosis, patient's (or deceased's) address at time of diagnosis, demographic factors (sex, race, date of birth, marital status, occupation, etc.), and risk factors (smoking history and toxic exposure if available). Verification sources include the MCR database, CHIME hospital discharge and mortality data, and physician records.

Case Ascertainment

Cases are ascertained if, based on the results of a Level 2 Inquiry, a Level 3 Inquiry is recommended by the Chronic Disease Medical Epidemiologist or the CI Committee. Case ascertainment involves an active search for additional cases in the study area, either by CI staff or Office of Surveillance, Research, and Evaluation (OSRE) and MCR staff. It also includes double-checking for inconsistencies in diagnosis, date of diagnosis, date of death, and additional verification of address, demographics, time residing in the study area, occupational activities, and geocoding for precise determination of the geopolitical study area.

The search for additional cases and cross-checking of case information is implemented by certified tumor registrars (CTRs) from OSRE-MCR in collaboration with registrars at reporting facilities in the study and surrounding areas. OSRE-MCR staff also implement crosschecking by matching cases in the Patient Listing Forms with cases in MCR, hospital discharge, and mortality databases. The geocoding is obtained by utilizing a geographical software (ArcView) and an updated census database (NCHS) to match addresses of cases on the Patient Listing Forms as well as those in the additional search to the specific geopolitical location of the study population.

Methods for Detecting Space and Time Clustering

Methods for detecting space and time clustering may be either proactive or reactive. Proactive methods involve identification of potential clusters during systematic and periodic cancer surveillance activities. Reactive methods involve the detection of space and time clustering after suspicion of a cluster is reported. The rate ratios described below and the Texas method are appropriate for proactive surveillance. All methods described in this section can be used for reactive cancer surveillance.

Rates, Ratios, and Confidence Intervals

Whenever possible, calculation of standardized rates and ratios will be the methods of choice to determine space and time clustering of cancer cases.

At both Level 2 and Level 3 of a cancer inquiry, whenever data is appropriate, incidence and mortality rates and the corresponding 95% confidence intervals will be calculated for the study population (as defined by study area and period of time). Direct or indirect methods for adjustment to different age distributions between comparison populations will be used.[5,6,7]

Rates and ratios will be calculated when there are at least 30 cancer cases arising from a large enough population. For directly standardized rates, the population denominator should be at least 50,000 persons. For indirectly standardized rates, the population denominator should be at least 10,000 persons.[2,8]

Because the standardized ratio is merely a weighted average of the age-specific relative rates (ratio of the index to the standard rate for each age group), the examination of age-specific relative rates should be implemented before any attempt is made to standardize. If the age-specific relative rates do not differ by sampling error, but instead vary systematically with age (e.g., either increase or decrease with age), no method of standardization should be applied, and comparison of the index with the standard population should be made within each age group separately.

In situations when study area age-specific (or other variable stratum-specific) rates are not available, only indirect methods can be used.

The standardized incidence and mortality ratios of the index (study) by the comparison population (SEER-US or MCR-Missouri) and the 95% confidence interval are used to evaluate space and time clustering when the number of cases and the denominator are sufficient to generate reliable rate and ratio estimates.

Direct Method

Using the direct method, the following steps should be implemented to obtain standardized rates and ratios:

1. Define a standard population (US 1970 Census Standard Population [Million] or the Missouri updated inter-census standard million population).
2. Apply the age-specific (or other characteristic) rates of the index population (study population) to the numbers in each age group of the standard population to obtain the number of cases that would be expected in the standard population if the index rate is applied.
3. Add the expected cases over the age group to obtain the total number of expected cases.
4. Divide the total in the expected cases by the total in the standard population to obtain the directly standardized incidence rate (DSIR) and the directly standardized mortality rate (DSMR) for the index (study) population.

The standard population crude rate is also, by definition, a standardized rate (the crude rate standardized to the population from which the rate is estimated). The ratio of the "directly" standardized rate of the index (study) by the standard (comparison) population is the comparative rate index (comparative incidence index [CII]; comparative mortality index [CMI]). Standard errors and accompanying 95% confidence intervals in both rates and ratios will use approximate methods.[7]

Also, the ratio of comparative indices in two communities, estimated using the same standard population (CII or CMI), equals the ratio of two directly standardized rates. In other words, directly standardized rates of two similar communities can be compared using either standardized ratios or rates (i.e., the method is consistent).

Indirect Method

Using the indirect method, the following steps should be used to obtain indirectly standardized rates and ratios:

1. Define a set of age-specific (or other characteristic) rates from the comparison population (i.e., the rates obtained from SEER-US or MCR-Missouri in this case).
2. Apply the comparison age-specific rates to the index (study) population to get the expected number of cases in each age group.
3. Add the expected cases over the age groups to obtain the total number of expected cases.
4. Divide the total of index (observed) cases by the total of expected cases to get the standardized incidence ratio (SIR) or standardized mortality ratio (SMR)

5. Multiply the crude rate in the comparison (Missouri) population by the standardized ratio to obtain the indirectly standardized incidence and mortality rates (ISIR and ISMR, respectively) in the index (study) population.

The ratio of index cases (observed) by expected cases is the "indirectly" standardized incidence or mortality ratio (SIR or SMR). Standard errors and accompanying 95% confidence intervals in both rates and ratios will use approximate methods.[7]

It is important to mention that in the indirect method the true standard is the index population and not the comparison populations from which rates were used to generate the expected number of cases in the index population. Therefore, the indirectly standardized ratios (SIR and SMR) of similar communities, though using the same set of comparison rates, cannot be directly compared because they have different standards (each community's own population).

Incidence or mortality ratios close to unity are an indication of no significant excess of cancer in the study area. Incidence and mortality ratios greater than unity for which confidence intervals do not include unity (null hypothesis) are considered an indication of cancer excess in the study population. The calculation of multiple SMRs across the sub-population inflates the type-I error when testing the null hypothesis of unity of SMR. A departure from the null hypothesis is considered when the standardized ratio confidence interval corresponds to p-values < 0.025.

Texas Method

The Texas method was designed to initiate *alert* and *action* levels of response to time clustering based on monitoring of data received by a surveillance system.[9] In proactive cancer surveillance, this method will be used in conjunction with indirect methods for rate and ratio calculations (SIR and SMR) to evaluate the evidence of increases in cancer incidence or mortality over time for a study area. Because cancer cluster evaluations are usually retrospective, the Texas method will also be used to evaluate these sporadic aggregations of cancer in reactive surveillance (e.g. community reports of clusters).

In the Texas method, the disease experience of the study area may be monitored by an SIR (or SMR) or Z-statistic corresponding to action and alert levels:

$$\text{SIR (or SMR)} = 1 + Z_1 / \sqrt{(E)}.$$
$$\text{SIR (or SMR)} = 1 + Z_2 / \sqrt{(E)}.$$

Where (E) is the expected value.

If the observed SIR (or SMR) or Z-statistic exceeds a pre-specified value, then an action level is reached. If the observed value falls below the action level, but above a pre-specified warning level, then an alert level is indicated. An action level is also signaled if two consecutive alert intervals are identified. The alert and action level can be expressed in terms of Z-statistic (Z_2 and Z_1) or standardized ratios (SMR_1 and SMR_2).

In order to determine the Z-statistic and the standardized ratio values, two parameters need to be specified. First is the probability that the action level will be exceeded in the first of two consecutive intervals (P_2). Second is the probability that the action threshold will be exceeded in either of the two consecutive intervals (P_3). Then, the probability of an alert will be:

$$P_1 = (P_2)^2 - 2P_2 + P_3$$

After evaluation of this probability, the one-tailed Z-scores associated with P_1 and P_2 are found from the table of probability by J. White, A. Yeats, and G. Skipworth.[10] These Z-scores represent the critical values for the alert and action mode levels. If the number of observed cases is small (i.e., less than 30), then a Poisson distribution may be assumed, and the critical values for alert and action levels may be interpreted as standardized ratios (SIR and SMR).

Additional Methods for Detecting Space and Time Clustering

The use of additional statistical evaluation of space-time clustering is appropriate for reactive surveillance, particularly under certain circumstances. First, when at any inquiry level, the number of cases is very small (i.e., less than 10 cases for most situations) or the denominator data is not available for calculating meaningful rates and ratios (i.e., population estimates by sub-population are not available or too small). Second, when at a Level 3 Inquiry, the consistency among the four space-time clustering statistical methods and the likelihood of observing the identified cluster are used to support a decision to initiate a Level 4 Inquiry.

In most situations, the space-time cluster statistical methods require a case series listed over time and either a comparison population from which an expected rate can be generated (e.g., Poisson method) or a background rate for the study area (e.g., Chen method).[1]

Poisson

In the absence of cancer cases clustering (a random occurrence) and because cancer occurrence is rare (small number of cases relative to the size of the population giving rise to cases), it

is assumed that the distribution of cancer cases over time in study and comparison populations follows a Poisson distribution.[1] In this framework, the job of statistical testing is to determine if the pattern of occurrence observed is a departure from a Poisson distribution and compatible with clustering of cases over time.

The Poisson distribution has only one parameter, μ, which is both its mean and variance. The occurrence of cancer cases over time t follows a Poisson distribution with expected number of cases=λt=μ and var=μ. The probability function of the Poisson distribution is the following:

$$P(X=x) = \mu^x e^{-\mu} / x! \quad x = 0, 1, 2 \ldots$$

The cancer cases reported (incidence and mortality) for the index (study) population during a specified period of time will compose the "observed" cancer cases. The observed cases may be generated by gender, race, and age-specific groups depending on the specificity of the inquiry. This specificity is based on the knowledge about the putative exposure and cancer occurrence relationship and mode of exposure. For childhood cancer, no further classification by gender is implemented.

The expected number of cases (μ) in the study population can be estimated by multiplying the rate in the comparison population (SEER-US or MCR-Missouri) during a given time period by the population at risk in the study area (Indirect Standardization).

Using these facts, the Chronic Disease Medical Epidemiologist can calculate the probability of observing **x** or more number of cases in the study population (over the time period) given that it is assumed the case occurrence follows a Poisson distribution for which the best estimate of its parameter is $\bar{x}$, the mean number of cases (for only one time period the mean number of cases μ equals the number of expected cases **x'**).

$$P(X \geq x) = \sum_{x=x}^{x=\infty} \mu^x e^{-\mu} / x! \qquad x = 0, 1, 2 \ldots$$

Conversely, one can calculate:

$$1 - [P(X \geq x) = 1 - \sum_{x=0}^{x=x-1} \mu^x e^{-\mu} / x!] \qquad x = 0, 1, 2 \ldots$$

In other words, what is the probability of observing x or more number of cases in the study population when the expected number of cases in the state is μ? It is assumed that when this probability is smaller than α=0.001, there is an excessive number of observed cases.

Chen

The Chen method is designed to detect only temporal aggregates of disease.[1] It compares the length of the observed time interval between successive cases with a critical interval based on the background rate of disease and the size of the population at risk. If each of the observed time intervals is shorter than the critical interval, a significant increase is determined.

For this method, a series of cases observed over many time periods is listed by date of occurrence. This list may be further classified into gender, race, and age-specific groups, if data is available. The observed length of time elapsing between the occurrence of consecutive cases is the variable of interest.

The expected time interval between successive cases in the study population during the study period is:

Expected Time Interval (ETI) = 12 * (1/ expected number of cases per study period)

In order to determine the number of cases per study period, one must first know the prior, background rate (incidence or mortality) of the cancer of concern in the study area. Then, multiply this rate by the population of the community estimated in the middle of the study period.

The critical interval is found by multiplying the expected time interval (ETI) between cases by a constant K defined as follows:

$$K = -\log (1 - p_0 (n)^{1/n}) ;$$

where: log is the natural logarithm to the base e.
p_0 (n) is the probability of type I error (e.g., 0.05).
n is the number of intervals between the first and the last case within the total period of the study.

A significant value is when each of the n observed intervals is shorter than the critical interval estimated (i.e., ETI * K).

Knox

The Knox method is useful to detect clustering occurring in both space and time.[1] All combinations of two cases and each of the two possible pairs are identified according to their spatial and temporal proximity to one another. Each case is identified in the X-Y coordinate plane. The critical time and space intervals are used to determine the temporal and spatial proximity. The latency of disease of interest and the population at risk will determine these critical intervals. Therefore, a total of N (N-1)/2 pairs, clustered in space and time, can be represented in a two-by-two table: Cell A denoting case-pairs that are close in space and time, cell B denoting case-pairs close in space, cell C representing case-pairs close in time, and cell D indicating case-pairs neither close in time nor space. The column totals are M1 and M2, denoting totals for pairs close in time and not close in time, respectively. Row totals N1 and N2 indicate totals for pairs close and not close in space, respectively. If one continues to adjust these two measures of closeness until N1/N and M1/N are nearly equal to a prescribed fraction (e.g. 0.10), then the expected value of A is (0.1) (0.1) N = 0.01N. It follows that:

Expected Value of A (m) = (N1)(M1)/N
A = x is the observed value of A

Then the probability of observing at least x cases given the expected number (m) is

$$P(X \geq x) = \sum_{X=x}^{X=\infty} m^x e^{-m} / x! \qquad x = 0, 1, 2 \ldots$$

or

$$1 - P(X \geq x) = \sum_{X=0}^{X=x-1} m^x e^{-m} / x!$$

The Barton method is adequate to detect changes in spatial patterns over time.[1] It utilizes an approach similar to analysis of variance.

If (f) is the period in which all cases occur and (n) is the total number of cases in this period, then the average interval length (d) between successive cases is:

$d = f / (n+1)$, for cases observed over the time period.

When the first case and the last case mark the beginning and the end of the period, then the interval is:

$d = f / (n-1)$

If the time interval between cases in a series is less than the average interval (d), then temporal clustering is present. There are (h) clusters of cases aggregated in a time interval and each time interval is a "time cell". The x-y coordinate plane is used to define a coordinate for each case using a grid. In this situation, an interaction between time and space is defined when cases close in space also tend to be close in time. If there is no interaction between space and time, then one expects that the average squared distance between the "centroids" of cases within each temporal cell and the overall centroid would be similar to the average squared distance between all cases in the sample (D) and the overall centroid.

The statistics of interest are:

Q = ratio of squared distances among (h) temporal cells to the mean squared distance between all cases (D);
V (Q) = Variance of Q; and $\sqrt{V(Q)} = Sd(Q)$.

A transformation of Q follows the F distribution. When the number of h temporal clusters are >3*Sd (Q) *N cases + 1, then Q follows a normal distribution. Then:

$Z = (Q-1) / Sd(Q)$.

Pearson

The Pearson method is used to detect spatial clustering.[9] It determines the probability that a distribution of cases within spatial cells is consistent with the null expectation that cases are randomly distributed among cells of equal size determined by population count.

Cells of same population size are defined by a geographical grid superimposed over a study area or by other convenient parameters such as a census tract, census block, or household.

It follows that with (n) cases and (m) cells, the number of cases per cell of interest is denoted by (s) and the observed number of cells with (s) cases is denoted by $(P)_s$.

Therefore:

The expected value $(\overline{P}_s) = m \binom{n}{s} p^s q^{(n-s)}$; s = 0, 1, 2...., n

Where p = 1 / m and q = 1-p.

The probability of observing a particular distribution of (n) cases among (m) cells is then estimated from the likelihood chi-square test:

$X^2 = \sum_s r_s^2$, where $r_s^2 = [\log (P_s / \overline{P}_s) + (\overline{P}_s / P_s) - 1]$; DF=K-1, where (K) is the number of different cell occupancy configurations observed.

Grimson

The Grimson method evaluates the evidence of clustering among adjacent areas (e.g. counties, zip codes, etc.).[10] It determines whether high-risk areas cluster spatially within a larger area. It compares the observed number of adjacent borders shared among high-risk areas with an expected number. The expected number of adjacent borders assumes that high risk areas are randomly distributed within the study region.

Give a maximum of N (N-1) / 2 pairs of areas, the probability of K adjacent counties follows a Poisson distribution

$P (X=x) = \mu^x e^{-\mu} / x!$ x = 0, 1, 2 ...

Then the important statistics are:

Mean of (K) = Expected value (K)= $\mu n (n-1) / 2 (N-1)$; and Var (K). If n is small and N is large, then K follows a Poisson distribution and $E(K) = V(K) = \mu$. However, if n / N is large, the probability of K follows a normal distribution.

Ohno

The Ohno method is designed to detect geographical aggregation of cases.[1] If adjacent areas experience levels of disease which are "more alike" than would be expected by chance, then clustering is present.

If a study area is partitioned into meaningful geographical boundaries, there are N cells and a total of "A" adjacent cell pairs. Also, if the rate of disease in each cell is calculated and rates are categorized into two or more levels, there are n_i cells in each rate category.

For N cells there are a total of N (N-1) / 2 possible disease category pairs, $n_i (n_i-1) / 2$ possible like-pairs within a category sub-set.

Therefore, the total number of adjacent pairs observed is compared to the expected number of adjacent pairs:

$$(A / N (N-1)) (n_i (n_i-1) / 2)$$

and to the expected number of like-pairs:

$$(A / N (N-1)) (\sum n_i (n_i-1) / 2)$$

The variable observed minus expected will follow a chi-square distribution with one degree of freedom.

REMSA

The Remsa method uses person, place, and time dimensions to determine whether particular occurrences tend to be non-randomly distributed within certain sub-groups defined by demographics and/or geographical location and/or time.[1] This method is useful when disease etiology is unknown and one needs to explore leads about disease risk factors.

The underlying study population is partitioned into study attributes (e.g. sex, race, place, month / year of diagnosis or death) and the proportion of the population that falls in each category is determined. Assuming independence among category-specific proportions, one can indirectly estimate the proportion of the population that falls in a sub-category determined jointly by two or more attributes by multiplying category-specific proportions. Population-based sample surveys can also be used to directly estimate these sub-category specific proportions.

Given N total cases distributed among all of the cells, and the probability of selecting a given cell equal to p, the probability of observing at least x of the N cases in that cell follows a binomial distribution:

$$P(X \geq x) = \sum_{x=x}^{x=n} \binom{n}{x} p^x (1-p)^{(n-x)} \qquad x = 0, 1, 2 \ldots$$

Limitations of Methods

General

Observations brought to the attention of public health officials are by definition "unusual" and therefore a "non-a priori situation." In this instance, statistical significance has little meaning.

Another issue is the many more than expected occurrences of cancer clusters, observed when cancer cases are distributed in an ever-increasing number of arbitrarily defined study areas. In the words of Bender "if an area is examined in detail for a long period of time, a statistically significant excess of cancer cases will be observed."[11] Clusters are not rare occurrences. With 250 million people in the United Sates and many types of cancer, chance alone may explain many identified clusters. Clusters occur continually within any large population, and their population occurrence is often no greater than that expected by chance alone. Therefore, cancer clusters often represent "expectedly unexpected events."[11]

Also, at both Level 2 and Level 3 of the CI process, many statistical tests may be implemented in the same population, constituting what is called multiplicity of statistical testing. When multiplicity is present, the likelihood of making a type-I error is much greater than that occurring by chance alone. In order to minimize this problem, the recommendation is to use a stricter level of acceptance for making a type-I error.[12]

In addition, the manner by which cancers are detected can affect rate comparisons. Screening and early diagnosis may change over time and by region. The rate comparisons and the statistical testing utilized assume this bias is not operating. Therefore, these variations may result in spurious differences when comparing rates and counts by time periods and regions.[13]

Both screening and diagnostic technique changes that are differential across regions and time periods can affect cancer staging. Distant stage (metastatie) cancer at some point in time may have been misclassified as localized or regionalized cancer at an earlier date due to this bias.[13]

Rate Standardization and Texas Method

In the case of a small-size index population, direct method estimates are unreliable. The variance of both directly standardized rates and ratios tends to be large compared to that estimated using the indirect method. This unreliability of the direct method may affect decisions regarding evidence for clustering based on the size of the ratios and the 95% confidence interval.

When rates are calculated using the direct method, rate differences and ratios, confidence intervals, and statistical testing are inaccurate if fewer than 10 cases are used to generate rates. No rate comparisons should be attempted if fewer than 6 cases are available.[14]

The application of the Texas method requires knowledge of the distribution of the index (study) population. For small communities, this information is not always readily available.

Both the indirect method for standardization and the Texas method suffer from lack of consistency when making comparisons among communities.

Also, because indirectly standardized ratios (SIRs and SMRs) are based on the age (or other factor) distribution of the index (study) population (rather than the standard), they are prone to change if there is a shift in the age (or other factor) distribution of this population over time. Therefore, comparison of SIRs or SMRs over time may, in effect, represent "different populations" as a result of demographic shifts.

Poisson and Chen Methods

The application of the Chen method requires knowledge of the population distribution and the baseline rate of the cancer of concern in the study area. Many cancer inquiries originate from small communities, for which expected rates are difficult to determine. The rates in these small communities are often unreliable, leading to substantial error in estimating expected rates and therefore expected time intervals.

Another limitation of the Chen method and also of the Poisson method is that these measures are not conventional parametric statistics. Therefore, a more conservative approach is recommended to prevent an increase in type-I error. The recommended p-value for decision criteria is $p<=0.001$ or smaller.[9]

Knox and Barton Methods

The main limitation of the Knox method is the dependency of results on the critical space-time distances (intervals) specified. This dependency may lead to considerable data "fishing" by retesting after changing of space-time critical distances. Both methods may also lose power by using the actual distance between case-pairs in the geographic coordinates. However, the use of geographical coordinates for each case in a continuous fashion allows the Barton method to overcome the subjectivity of determining spatial proximity present in the Knox method. A limitation of both the Knox and Barton methods is the reliance on the criterion for determining statistical significance on space-time interaction. If clustering is present in one dimension alone, either space

or time, the test statistics are likely not significant. However, a major strength of both methods is that they do not require prior knowledge of the size of the underlying population, values of demographics, and baseline frequencies of disease occurrence or rates.

Pearson Method

A major weakness of the Pearson method is the assumption that the population at risk is the same in each of the probability cells. In most situations, this assumption does not hold. Also, because the cells are frequently arbitrarily defined, observation of clustering within a set of cells may have little practical significance. This can be overcome, however, by exploring different cell configurations (e.g. census tracts, blocks, etc.).

Grimsom and Ohno Methods

A major limitation of both the Grimson and the Ohno methods is that counties (zip codes or census tracts) sharing a few miles of border are considered equivalent to counties sharing dozen of miles. In addition, in both methods, counties that have many borders have a greater likelihood of random clustering compared to those with only one or two borders. Nonetheless, tailor-made probability distribution is a strength of the Grimson method because it eliminates concern regarding appropriateness of alternative distributions. Although the Ohno method uses an alternative distribution, the chi-square, its validity has been demonstrated through Monte Carlo simulation.

REMSA Method

This method relies on the multiplication of probabilities for independent events. If attributes of the study population being evaluated are not independent, the occurence of events will not be independent. This situation is likely to occur when evaluating clustering by person, place, and time. For example, if the study population is 50% African American and 80% high socioeconomic status (SES), the REMSA method assumes the population of African Americans who are high SES to be 40%. It is possible, however, that the high SES and African-American population composes only 10-20% of the study population.

Epidemiological Translation

As mentioned in the section titled "Epidemiological Evaluation of Clustering," application of epidemiological criteria for disease causation and reasoning should guide the CI process. Statistical tools are used to assist in making a decision about the existence of a cluster, never to determine this decision.[5]

The criteria for causality of an association that should guide the epidemiologist, and therefore the cancer inquiry process, are the consistency, biologic plausibility, and evidence of dose-response.

The consistency criterion requires a wide knowledge base of the natural history and descriptive epidemiology of the reported cancer(s), as well as the use of a systematic literature review that incorporates the principles of evidence-based research. In terms of cluster investigation, consistency may mean one of the following:[1]

- historical patterns in the reported cases
- pattern of occurrence consistent across reported literature
- homogeneity of reported cases (e.g. same sex, race, age, or occupation; same cell type, anatomic sites, or pathway of exposure)
- consistency within aggregation (within cluster).

Biologic plausibility requires a strong knowledge of the disease process of concern and reliable information on exposure of interest. In the context of a cancer inquiry, the biologic plausibility may mean one of the following:

- presence of an environmental or occupational risk (e.g. asbestos abatement and mesothelioma)
- demonstration that a pathway for exposure is possible
- recognition that a specific organ or tissue is a possible site for biologic action of the suspected exposure.

In analytical epidemiology and medical research, dose-response is considered by far one of the strongest criteria for determining a causal association. However, cluster investigations are hypothesis-generating types of research, as opposed to analytical studies that are designed to test hypotheses of causation. This difference should be carried over to the careful application of criteria and the interpretation of the findings of a cluster investigation. Often the dose-response is evaluated against the time continuum. For this reason, a series of reported cases over a long period of time is usually required to assess this criterion in a meaningful way. Sometimes dispersion of spatial patterns may also be used to assess dose-response.

In a cluster investigation, dose-response usually means:

- duration of exposure (e.g., a greater proportion of cases with long-term residency in the proximity of the hazard or proximity in time of suspected events).

In a cluster investigation, dose-response less often means:

- identification of a specific spatial pattern of dispersion (e.g., the closer the proximity to the exposure, the greater the number of observed cases); or,
- a combination of duration of exposure and spatial pattern of dispersion.

Abbreviations

BCC	Bureau of Cancer Control
CHIME	Center for Health Information Management and Epidemiology
CI	Cancer Inquiry
CII	Comparative Incidence Index
CMI	Comparative Mortality Index
CTR	Certified Tumor Registry
DNR	Department of Natural Resources
DOA	Department of Agriculture
DOH	Department of Health
DSIR	Directly Standardized Incidence Rate
DSMR	Directly Standardized Mortality Rate
ISIR	Indirectly Standardized Incidence Rate
ISMR	Indirectly Standardized Mortality Rate
MCR	Missouri Cancer Registry
NCHS	National Center for Health Statistics
NIOSH	National Institute of Ocupational Safety and Health
OSHA	Occupational Safety and Health Administration
OSRE	Office of Surveillance, Research and Evaluation
SEER	Surveillance, Epidemiology and End Results Program
SEPH	Section for Environmental Public Health
SIR	Standardized Incidence Ratio
SMA	Standardized Metropolitan Area
SMR	Standardized Mortality Ratio

References

1 Cluster 3.1 Software System for Epidemiological Analysis, Instruction Manual, Centers for Disease Control and Prevention, Agency for Toxic Substances and Disease Registry, US Department of Health and Human Services, Atlanta, GA 1993.

2 Kleinman, Joel C, "Age-Adjusted Mortality Indexes for Small Areas: Applications to Health Planning" *American Journal of Public Health*, 67 September 1977: 834-840.

3 Gloecker Ries, Lynn A., et al, eds. *SEER Cancer Statistics Review, 1973-1994*. U.S. National Cancer Institute, 1997.

4 Schmidt KM, Aaron, SW. Illustrated ICD-9-CM Code Book. Reston, VA: St. Anthony Publishing, 1996.

5 Rothman KJ. Standardization of Rates. In KJ Rothman (eds.), Modern Epidemiology (1rst ed.). Boston/Toronto: Little, Brown and Company,1986.

6 Osborn JF. Vital Statistics: Standardization. In JF Osborn, Manual of Medical Statistics, Volume 1, London School of Hygiene and Tropical Medicine, Division of Medical Statistics and Epidemiology, 1986-87 edition.

7 Kahn HA and Sempos CT. Adjustment of Data Without Use of Multivariate Models. In HA Kahn and CT Sempos (eds). Statistical Methods in Epidemiology, Monographs in Epidemiology and Biostatistics, Volume 12 (rev. ed. of: An Introduction to Epidemiology Methods/Harold A Kahn.1983). New York/Oxford: Oxford University Press. 1989).

8 Venturea, Stephanie J, et al. "Births and Deaths: United States, 1996" *Monthly Vital Statistics Report*, 46 September 11, 1997

9 Hardy RJ, Buffler P, Prichard HM et al. Monitoring for Health Effects of Low-Level Radioactive Waste Disposal: A Feasibility Study, Texas Department of Health, 1983 In: Cluster 3.1 Software System for Epidemiological Analysis, Instruction Manual, Centers for Disease Control and Preven tion, Agency for Toxic Substances and Disease Registry, US Department of Health and Human Services, Atlanta, GA 1993.

10 White J, Yeats A, Skipworth G. Tables for Statisticians (Third Edition) Hong Kong: Stanley Thones Ltd., 1985.

11 Bender AP, Williams AN, Johnson RA, Jagger HG.. Appropriate Public Health Response to Clusters: The Art of Being Responsibly Responsive. Am J Epidemiol.1990; vol 32(suppl.):S48-S51

12 Samules, ML. Statistics for the Life Sciences. San Francisco: Dellen Publishing Company 1989; pg. 500-504.

13 US Department of Health and Human Services, Public Health Services, National Institute of Health, Nation Cancer Institute, Bethesda, Maryland 20892, NIH Publication No. 89-2789.

14 Breslow, NE, Day, NE. Statistical Methods in Cancer Research, vol II, Lyon, France: IARC 1987, p. 59.

Appendix A

Example
Letter for Individual Cancer Concern

Dear Resident:

Thank you for taking the time recently to talk with my co-worker and to review with him your concerns about the occurrence of cancer among your family members.

As I understand it, your son-in-law, who is forty, has been diagnosed for the second time with Hodgkin's lymphoma, and his son (your grandson), who is fourteen, has been diagnosed with Burkitt's lymphoma. You have had the private well at their home checked by a county health official, and no contaminants were found. Your daughter is gathering information on the diseases affecting her husband and son. She has been able to find information regarding Hodgkin's lymphoma, but has not been able to find information regarding Burkitt's lymphoma. You would like any information that we can find on Burkitt's lymphoma.

In response to your concerns, we first contacted the Missouri Cancer Registry, which maintains a database of new cancer cases reported by Missouri hospitals. For the time period 1985-1992, the years for which the Registry has the most accurate data, 44 cases of Burkitt's lymphoma were diagnosed in Missouri hospitals.

Also in response to your concerns, Cancer Inquiry Project staff have researched the subject of Burkitt's lymphoma, using available books from our library as well as searching the Internet for articles from various medical journals. The results from our research are summarized below.

Lymphomas are broadly separated into two different categories: Hodgkin's and non-Hodgkin's lymphoma. Non-Hodgkin's lymphoma (NHL) occurs more frequently than Hodgkin's in children, and frequency increases throughout childhood. Boys are two to three times as likely to be affected than girls. While cancer of any type is less common in children, cancers of the lymph system are among the most common in childhood. Burkitt's and Burkitt's-like lymphomas are classified as undifferentiated (small noncleaved cell) B-cell lymphomas, and account for about one third of NHL in children. Childhood NHL is the type of cancer most commonly associated with inherited immune deficiencies.

The type of NHL known as Burkitt's lymphoma is triggered by a change in one of three specific parts of the chromosomes of the lymph cell. This change leads to uncontrolled growth of the new, malignant cell. For Burkitt's in particular, however, there is another strong risk factor besides (or in conjunction with) an inherited condition, which is exposure to a virus known as the Epstein-Barr virus (EBV). This is a virus to which most people in the world are exposed at one time or another, including even remote Amazon tribes. While there are usually no signs of sickness, it is the cause of mononucleosis, also known simply as

'mono', or 'the kissing disease'. Worldwide, exposure to this virus in conjunction with some other, as yet unknown, factor is believed to be responsible for most cases of Burkittís lymphoma. The association is not as strong, however, in the United States. Besides being suspected in many NHL cases, EBV is also a suspected risk factor for developing Hodgkin's lymphoma.

Dr. Eduardo Simoes, a consulting epidemiologist here at the Department of Health, noted that the information you reported to us did not include any mention of these diseases among neighbors or other nearby residents; it appears to be confined to a single household. The two diseases are occurring among two genetically similar individuals (father and son). This is known as familial clustering. It is unlikely, therefore, that any environmental or occupational exposures such as chemicals or radiation would be responsible. The most likely scenario appears to be that of two genetically predisposed individuals who were exposed to Epstein-Barr virus (or a similar virus), which triggered growth of lymphoma tumors.

While there is no doubt as to the seriousness of the diseases you reported to us, it may be encouraging to know that both Hodgkin's and Burkitt's lymphomas tend to respond very well to treatment, and have a very high survival rate.

I am also enclosing a booklet from the American Cancer Society titled Cancer Facts & Figures - 1997. This booklet provides general information about rates of cancer, plus more specific information about the suspected causes of several of the more common types or sites of cancer. You may note that the risk factors for the various types of cancer are often quite different from each other. For example, some of the other risk factors for non-Hodgkin's lymphoma include age, gender, and reduced immune function. Rates increase steadily with age. Males have higher rates than females. Persons with organ transplants are at higher risk due to altered immune function, as are persons infected with AIDS. Exposures to high-dose radiation and exposures to agricultural chemicals are also suspected risk factors. On the other hand, the most important risk factor for lung cancer is cigarette smoking. Certain industrial exposures, such as exposure to substances like asbestos, exposure to radiation, or exposure to radon can also increase the risk of lung cancer. The risks from asbestos and radon are multiplied by cigarette smoking. Exposure to environmental tobacco smoke or secondhand smoke elevates lung cancer risk slightly in nonsmokers. Lung cancer risk is also higher for people with a family history of lung cancer and for individuals who have been previously diagnosed with a nonmalignant lung disease, such as tuberculosis or asthma.

Your concern about cancer is understandable given the high rates of cancer and deaths from cancer that occur across the country, and considering your family's personal fight with the disease. Cancer is the second leading cause of death in the U.S. and Missouri. About one out of every three persons will develop cancer at some time during their life, and three out of four families will be affected.

Again, thank you for your time and concern. I hope this information is helpful to you. If you have any further questions about cancer, please contact us again in the near future at (573) 522-2845.

Sincerely,

Cancer Inquiry Coordinator

Appendix B

Inquiry number:

Date:

Cancer Inquiry
Initial Report Form

Name of Staff Member Completing this form:

Inquirer making inquiry:

Name:

Title or Group Represented:

Address:

City/State/ Zip:

Phone Number(s):

Best Time to Contact:

City/County of Concern:

Zip Codes of Concern:

Boundaries:

Types of Cancer of Concern:

Suspected Causes:

Level of Interest:

Other Public Officials Contacted:

Name:

Title:

Agency:

Address.

Phone Number:

Summary of Inquiry/ Telephone Conversation:

Appendix C

Initial Letter

Date
CI97-000

Mrs. XXXXXXXX
XXXXXXXXXX
XXXXXXXXXX

Dear Mrs. XXXXX,

Thank you for your recent call/letter concerning cancer in your neighborhood/ school/ county/ workplace. You have clearly put much thought into what has occurred among your family/ neighbors/ co-workers.

Cancer is a very serious and frightening, but surprisingly common disease. Current information shows that approximately one out of three Americans will develop cancer in our lifetimes. It is estimated that the disease will affect three out of four families. The chance of developing cancer also increases with age, so as our population ages we will see more cases of cancer in our communities.

It is important to realize that cancer is not one disease, but many. It can occur in almost all organs in the body. Each organ can have a different type of cancer that develops in it and each type has different risk factors. For example, the main risk factor for lung cancer is cigarette smoking, but for skin cancer it is sun exposure. On the other hand, it really isnít known what causes some cancers like breast cancer.

Many people believe that something in the environment causes the cancer they see in their community, but behavior accounts for most of our known cancer risk. Such factors as smoking, diet, heavy alcohol use, sexual and reproductive history can all contribute to our developing cancer. It is estimated that less than 10% of cancers are caused by environmental exposures. In contrast about 30% of cancer is probably caused by cigarette smoking alone. In addition, family history is important in many cancer cases. Even though thousands of studies have been done about the causes of cancer, we still do not know exactly what causes most cancer cases.

Most cancers take a long time to develop. It is usually decades from the time someone is exposed to something that might cause cancer to the time that they find out they have cancer. This is one of the reasons that cancer is more common as we get older. In general, cancer is usually caused by something that happened five to forty years ago. In addition, the few chemicals that we think are likely causes of cancer must have fairly long and concentrated exposures before they will cause cancer. This level of exposure is usually found in work settings, not our homes.

We call an apparent increase of cancer in one community or group a cancer cluster. If a cluster includes cancers of different types, it is probably not a "true" cluster. For example if someone called us and reported that there were many people with cancer in their community, but the kinds of cancer included lung, breast, leukemia, and prostate, we would usually determine that that was not a true cluster. As you can see, in order for us to conduct a study, it is important for us to know the details about the type(s) of cancer which seem to be part of a cluster.

The Department of Healthís Cancer Inquiry Program, however, does have a procedure to examine potential cancer clusters. A ìtrueî cancer cluster that we can study usually has one or more of the following characteristics:

there is only one type of cancer involved

the cancer is occurring in an age group that we would not expect

there is a very rare type of cancer involved

there is a scientifically established causal relationship between the type of cancer and the suspected exposure.

In addition, for a scientific study to be conducted, we must have hundreds of cases in a large area. Most of our knowledge about cancer causes comes from studies comparing many people with the same cancer to people who do not have cancer. Investigation of a few people with different cancers will not shed light on the cause of cancer because those cancers were probably caused by different factors.

The fact that cancer is common does not mean that you, your neighbors, or the Missouri Department of Health should not be concerned. Many of these cancers are preventable. There are national and state goals to lower the number of people with cancer in our communities. In order to do that, we need people who are interested in the problem and willing to help raise awareness about how to prevent cancer.

I am enclosing several fact sheets and brochures to help you understand more about cancer. Please read them and if you would like to initiate a preliminary investigation, please fill out the enclosed forms as completely as possible and return them. It is important that we get these forms returned, so just provide the best information you have available. If we do not hear from you in six weeks, we will assume that you do not want us to pursue this further. If you feel that you need more than six weeks to gather the information, or you have any questions, I can be reached at ((573) 522-2845.

Sincerely,

Cancer Inquiry Coordinator

enclosures

c:

Appendix D

Cancer Excess Factsheet

Cancer inquiries

The concerns regarding a possible excess of cancer usually occur when someone's spouse, neighbor, or co-worker is diagnosed with cancer. This close contact with cancer often brings an awareness of others who have cancer and may lead to the perception that there is an unusually large number of neighbors or co-workers with cancer. It is not uncommon for people to suspect the cancer cause is some toxic substance in the environment. Increased awareness about cancer and the search for a cancer cause may lead someone to contact the Missouri Department of Health or their local health department. The following cancer facts might help answer some questions about cancer.

Cancer facts

Cancer is more common than most people realize.

According to the American Cancer Society, about one out of three Americans now living will eventually have cancer. Over the years, cancer will strike about three out of four families. As science and medicine have conquered many diseases which used to contribute to premature death, cancer is now the second leading cause of death in the United States, following heart disease. Given these statistics, it is not surprising to know several people in a neighborhood or workplace who have cancer.

Cancer is not just one disease.

Cancers are a group of more than 100 diseases characterized by uncontrolled growth and spread of abnormal cells. Different types of cancers have differing rates of occurrence, causes, and chances for survival. Therefore, we cannot assume that all the different types of cancers in a community or workplace share a common cause.

The risk of having cancer is related to age.

While cancers occur in people of all ages, incidence rates for most types of cancers rise sharply among people over 45 years of age. When a community, neighborhood, or workplace consists primarily of people over the age of 45 and, even more so over the age of 60, we would expect more cancers than in a neighborhood or workplace of diverse ages. However, cancer is also the second leading cause of death in children (the leading cause is accidents).

Most cancers are related to lifestyle factors.

Cancers may be caused by a variety of factors acting alone or together, usually over a period of many years. Scientists estimate that most cancers are due to factors related to how we live. Lifestyle factors which increase the risk for cancer include: smoking cigarettes, drinking heavily, and diet (for example, excess calories, high fat, and low fiber). Other important cancer risk factors include reproductive patterns, sexual behavior, and sunlight exposure. A family history of cancer may also increase a person's chances of developing cancer.

Toxic substances in the environment account for a relatively small percentage of all cancer deaths.

Many people believe that cancer is usually caused by exposure to toxic substances in the home, community, or workplace. Although we do not know the exact impact now of environmental pollutants on cancer development, less than 10% of cancers are estimated to be related to hazardous environmental exposures.

Cancers today are usually related to events that happened many years ago.

For those instances in which a cancer is due to a contact with a cancer-causing agent, the disease does not develop immediately. Instead, there is often a long period, 10 to 30 years, between the exposure to a carcinogen (a cancer-causing substance), and medical diagnosis of cancer. This makes it very difficult to pinpoint what caused the cancer because the cancers we see now are usually related to a lifetime of certain habits or exposures to carcinogens many years ago.

Cancer clusters can occur by chance.

Determining that a true "cluster" or elevated number of cases exists requires a very sophisticated statistical analysis. We usually cannot conclude that this increase was caused by exposure to a particular environmental carcinogen. In addition, problems of statistical inference based on small numbers of cancer cases usually mean that statistical analyses would not yield useful or valid information

Experience in the United States with cancer cluster inquiries and investigations.

Since the 1970s when the state cancer registries were first being organized, many public health scientists and citizens hoped that anecdotal observations of clusters of cancer in the community might lead to prevention of new cases via the discovery of specific causes of these cancers. Since then, thousands of investigations have taken place throughout the country, mainly conducted by state, local, or federal agencies. With one or two possible exceptions involving childhood cancers, none of these investigations have lead to the identification of the causes of any of these possible clusters, even when a statistical elevated number of cancers in a geographic area could be documented. At the Missouri Department of Health, active surveillance of geographic patterns of cancer incidence is being conducted, made possible in part by electronic data reporting and through the Missouri Cancer Registry. We hope that systematic identification of cancer patterns in the population will lead to useful opportunities for prevention and control of cancer.

For more information, contact Stan Cowan by phone at 573-522-2845 or by mail at P.O. Box 570, Jefferson City, MO 65102-0570.

This fact sheet was adapted in part from a fact sheet prepared by the Wisconsin Department of Health and Social Services, and the New Jersey Department of Health and Senior Services.

Appendix E

Example Risk Factor Information

Risk Factors for Breast Cancer

From David Schottenfeld, Cancer Epidemiology and Prevention, Second Edition, 1996.

The main cause of breast cancer is thought to be estrogen. The more breast tissue is exposed to estrogen, the higher the risk that cancer will develop. The amount of estrogen in the body can be affected by many factors. As it relates to breast cancer, the most important factor in controlling the amount of estrogen in the body is the menstrual cycle. The more cycles a woman has in her life, the more estrogen is released into her system. Therefore women with early menarche (at or before 12 years of age) and/or late menopause (after 55 years of age) are more likely to develop breast cancer. Estrogen is most prevalent in the last stage of the cycle (after ovulation) which always occurs approximately two weeks before the beginning of the next cycle. Therefore, the shorter and more regular a woman's cycle, the higher her chance of developing breast cancer is.

Pregnancy also affects the supply of estrogen in the body and therefore on breast cancer. The older a woman is when she first gives birth, the greater her chance of developing breast cancer. Women who have never had children are at a higher risk of developing the disease. Abortion (either natural or induced) before the first birth probably increases a woman's chance of developing breast cancer. Women who do not nurse after they have a child also are at increased risk compared to women who do nurse, because nursing delays the re-establishment of regular menstrual cycles (and therefore the release of estrogen) after pregnancy.

Lifestyle factors also affect the risk of developing breast cancer. High levels of physical activity might delay menarche and, especially in young women, decrease the regularity of the cycle, thereby decreasing the number of cycles a woman has in her life. Therefor, low physical activity (especially in adolescents) might increase the risk of developing breast cancer. Because body fat releases estrogen, being overweight, especially after age 60 is a suspected risk factor for breast cancer.

Dietary factors might also play a role in the development of breast cancer. A diet high in fat and calories is a suspected risk factor. Case-control studies have been used to demonstrate this association. Cohort studies however found no evidence of a link between a high fat or high calorie diet and an increased risk of breast cancer. Animal studies and case-control studies have shown that a diet high in fiber may protect against breast cancer. More than three alcoholic drinks per day may increase the risk of developing breast cancer. This has been demonstrated through a large number of studies.

There is speculation that exposure to pesticides, especially organochloride and other halogenated compounds, whose effects mimic estrogen, might cause of breast cancer, but this has not been confirmed

Family history also has some influence on a woman's risk of developing breast cancer, but the link is not totally clear. It appears that some women have a genetic predisposition for breast cancer, but it seems possible that a genetic predisposition needs another trigger for a woman to develop cancer. BRCA1, as well as other chromosomal anomalies, have been linked to breast cancer.

A history of benign breast disease increases the risk of developing breast cancer. Chronic cystic or fibrocystic disease has a stronger association with breast cancer than fibroadenoma, but studies are not conclusive.

MISSOURI DEPARTMENT OF HEALTH
BUREAU OF CANCER CONTROL

LISTING OF CANCER PATIENTS

Please complete all information as completely as possible for each person. At a minimum, we must have a *legal* name, birthday, how long the patient lived in the area and where he or she was diagnosed in order to confirm this information and use it to look into your concern.

Name: _____ Sex: _____

Address: _____ Date of Birth: _____

_____ Date of Death: _____

Phone number: _____ Date of Diagnosis: _____

Type of Cancer: _____

Physician's Name: _____ Hospital: _____

Name of the facility where this person was first diagnosed with cancer: _____

Years at above address: _____

Comments: (Include maiden names or previous names, if any. Information about smoking status,

occupation, exposure, etc.): _____

_____

Name: _____ Sex: _____

Address: _____ Date of Birth: _____

_____ Date of Death: _____

Phone number: _____ Date of Diagnosis: _____

Type of Cancer: _____

Physician's Name: _____ Hospital: _____

Name of the facility where this person was first diagnosed with cancer: _____

Years at above address: _____

Comments: (Include maiden names or previous names, if any. Information about smoking status,

occupation, exposure, etc.): _____

_____

Return to: Comprehensive Cancer Control Coordinator
Missouri Department of Health
920 Wildwood, P. O. Box 570, Jefferson City MO 65102-0570
(573) 522-2843 fax (573)522-2899

MO 580-2276 (5-99)

MISSOURI DEPARTMENT OF HEALTH
BUREAU OF CANCER CONTROL

CANCER INQUIRY PROGRAM RESPONSE FORM

Name: _____

Address: _____

_____

Phone number and time we can reach you:_____

Please write a detailed description of what you suspect is causing the cancer in your community

(attach extra paper if necessary):_____

_____

_____

_____

_____

How do you believe people were exposed to this (through the air or water, on the job, etc.): _____

_____

Number of cancer cases that are on the patient listing forms:_____

Number of these that are deceased: _____

Number of types of cancer:_____

List the types of cancer (e.g. breast, colon, brain): _____

_____

Time period when these cases were diagnosed (e.g. 1989-1994):_____

Time period of exposure: _____

Is this exposure still occurring: _____

Have you spoken to any other government agencies about this possible exposure:_____

If yes, please list them and the person there who you spoke to: _____

_____

_____

CONTINUE ON BACK IF NECESSARY.

MO 580-2277 (4-98)

Appendix G

Letter Stating that forms were not returned

Dear Ms.,

The Cancer Inquiry (CI) Program Staff has been reviewing its files in preparation for its upcoming Cancer Inquiry Committee Meeting. A letter was sent Date, requesting your assistance in researching your cancer concerns. This review showed that the staff has not received the Response or Patient Listing Forms that were sent with the initial letter concerning your inquiry.

Because no further research into your concerns can be done without that information, if the CI Program does not hear from you within two weeks of the date of this letter, this inquiry will be considered closed. If you have any questions, please feel free to call me at phone number.

Sincerely,

Cancer Inquiry Coordinator

Date

Inquiry Number

Appendix H

Letter Confirming that the CI Program will do Further Research

Address
Address
Address
Address

Dear Ms. Name:

Thank you for returning the Response Form and Patient Listings that we sent you.

The next step in our process is to begin to check the list of cancer cases you sent us against our records to verify the information on the response form and the Patient Listings. We will then begin holding the information you provided us against a set of criteria to determine if we can do further research. If the information is complete enough for further research we will begin that process and let you know the results as soon as possible. Please be patient. It may take between a few weeks to a few months before we can complete our research. If it is determined that there is no clustering of cancer cases in your community/neighborhood/ occupational setting/school and the identified environmental exposure does not represent a community health hazard, you should receive a description of these findings within xxx weeks. On the other hand, if we confirm that there is a possible cancer cluster or that an environmental hazard is present in your community, further research and evaluation may take two to three months to complete, after which we will provide you a detailed description of our findings.

This inquiry will be discussed at the next Cancer Inquiry Committee meeting on ___*(date)*___.

In the meantime, if you have any questions or any new information that you think might be helpful, please feel free to call me at (573) 522-2845. Thank you for the concern you have shown about your community.

Sincerely,

Cancer Inquiry Coordinator